



Исследования и разработки в области новых информационных технологий и их приложений

Research and development in the field of new IT and their applications

УДК 004.422

DOI: 10.25559/SITITO.14.201801.170-182

БАЗИС ФОРМИРОВАНИЯ ПРОГРАММНО-МАТЕМАТИЧЕСКОГО ОБЕСПЕЧЕНИЯ В ЗАДАЧАХ ИТ ОБРАЗОВАНИЯ

Н.В. Зорина, В.М. Панченко

МИРЭА – Российский технологический университет, г. Москва, Россия

Аннотация

В статье рассмотрены проблемы и задачи разработки программно-математического обеспечения основных бизнес-процессов вуза на примере ИТ образования. Обоснована необходимость использования методов анализа при разработке программно-математического обеспечения для ИТ систем современных вузов и перечислен ряд актуальных задач, к решению которых можно подойти с помощью предлагаемого фреймворка. В работе описаны исследовательская гипотеза, цель, методология и этапы проведения исследований, а также достигнутые результаты. Материалы исследования представляют из себя априорные (ретроспективные) и апостериорные (текущие) образовательные данные. Эти данные получены из общедоступных источников и содержат сведения об учебной деятельности в виде результатов экспериментальных наблюдений по репрезентативной выборке учащихся. Для формального описания полученных данных использовано представление на основе математического аппарата теории множеств и алгебраических структур. Предложен авторский метод классификации выявленных источников образовательной информации по трем значимым признакам. Проведен анализ бизнес-процессов, отражающих взаимодействие учащихся между собой и взаимодействия учащегося и преподавателя в процессе обучения. На их основе предложена модифицированная модель архитектуры системы управления учебным процессом вуза. Данная модель построена на основе бизнес-процессов коллаборации и кооперации при осуществлении учебной деятельности. Она отражает изменения, которые произошли за

Об авторах:

Зорина Наталья Валентиновна, старший преподаватель, кафедра инструментального и прикладного программного обеспечения, МИРЭА – Российский технологический университет (119454, Россия, г. Москва, пр. Вернадского, д. 78); ORCID: <http://orcid.org/0000-0001-7893-9227>, zorina@mirea.ru

Панченко Виктор Михайлович, кандидат технических наук, профессор, кафедра инструментального и прикладного программного обеспечения, МИРЭА – Российский технологический университет (119454, Россия, г. Москва, пр. Вернадского, д. 78); ORCID: <http://orcid.org/0000-0002-7038-878X>, pvm36@yandex.ru

© Зорина Н.В., Панченко В.М., 2018



последние пять лет в связи с активным внедрением цифровых способов связи и интерактивного взаимодействия. Приведен перечень имеющихся инструментов для проведения разработки с использованием методов анализа данных, перечислены их достоинства и недостатки. Обоснован выбор инструментального средства, среды разработки и языка программирования для модуля анализа данных в составе фреймворка. Разработанный базис позволит подойти к задаче адаптации и персонализации учебного процесса с использованием современных ИТ-технологий и технологий Big Data, а также решить частично проблему унаследованных ИС отечественных вузов на основе использования накопленных данных.

Ключевые слова

Извлечение знаний; большие данные; анализ образовательных данных; учебная аналитика; извлечение знаний из образовательных данных; системный подход; программно-математическое обеспечение; инструментальные средства разработки; аналитическая платформа KNIME.

BASIS OF FORMATION OF SOFTWARE-MATHEMATICAL SUPPORT IN TASKS OF IT EDUCATION

Natalya V. Zorina, Victor M. Panchenko

MIREA - Russian Technological University, Moscow, Russia

Abstract

In the article problems and tasks of software development and mathematical support of the basic business processes of the university are considered on the example of IT education. The necessity of using analytical methods in the development of mathematical software for the IT systems of modern universities, it also lists a number of urgent tasks that can be addressed with the help of the proposed framework. The paper describes the research hypothesis, the purpose, methodology and stages of research, as well as the achieved results. The research material represents a priori (retrospective) and a posteriori (current) educational data. These data are obtained from publicly available sources and contain information on educational activities in the form of the results of experimental observations on a representative sample of students. For a formal description of the data obtained, a representation based on the mathematical apparatus of set theory and algebraic structures was used. An authorial method for classifying the identified sources of educational information on three significant grounds is proposed. The analysis of business processes reflecting the interaction of students among themselves and the interaction of the student and teacher in the learning process is carried out. A modified model of the architecture of the management system of the teaching process of the university is proposed on this business processes. This model is based on the basis of business processes of collaboration and cooperation during the implementation of educational activities. It reflects the changes that have been occurred in the past five years due to the active introduction of digital communication and interactive interaction. The list of available tools for development using data analysis methods is given, their advantages and disadvantages are listed. The choice of the tool, IDE and programming language to analyze the data module as part of the framework is substantiated. The developed basis will make it possible to approach the problem of adapting and personalizing the learning process using modern IT technologies and Big Data, as well as partially solve the problem of the inherited IS of Russian universities based on the use of collected data.



Keywords

Data Mining (DM); Big data; analysis of educational data; Learning; Learning Analytics (LA); Educational Data Mining (EDM); systems approach; software and software development tools; analytical platform KNIME.

Введение

ИТ образование в современном вузе России относится к быстро меняющемуся направлению подготовки кадров, что полностью укладывается в рамки философии Smart Educational и означает высокотехнологичное, умное или гибкое образование [1]. Анализ работ в области анализа данных в образовательных учреждениях [2-11] выявил два перспективных направления: учебная аналитика и анализ образовательных данных. Оба направления схожи тем, что используют образовательные данные, накапливаемые вузами, но в то же время они преследуют разные цели. Схожесть и различия были ранее описаны в работе авторов, посвященной интеллектуальному анализу результатов образовательной деятельности¹.

Разработка программно-математического обеспечения по автоматизации задач управления образовательным процессом для современного ИТ-образования неразрывно связана с применением самых последних достижений в области применения информационных технологий. Ряд задач, которые каждый день решают преподаватели и административный персонал учебных отделов подразделений вузов, связан с тем, как сделать образовательный процесс управляемым и улучшить результаты образовательной деятельности студентов.

Современные информационные системы должны строиться таким образом, чтобы использовать накапливаемые во время обучения данные: 1) для учебной аналитики, построения и визуализации статистической отчетности; 2) извлечения знаний и построения прогностических моделей; 3) выдачи на этой основе рекомендаций для всех пользователей

системы. Накапливаемые в информационных системах образовательные данные и учебная аналитика помогут сделать “видимыми” данные, которые до этого были невидимыми, незаметными и, следовательно, никак не использовались [2].

Использование программно-математического обеспечения, в основе которого лежат методы анализа образовательных данных, позволит подойти к решению следующих задач:

Задача поиска скрытых закономерностей, взаимозависимостей, дающих чёткое понимание того, как именно учащийся усваивает знания, приобретает умения и навыки;

Задача улучшения управляемости учебным процессом методами поддержки принятия рациональных решений;

Задача направления учащихся по наиболее целесообразному образовательному пути – помощь в выборе оптимальной образовательной траектории;

Задача генерации рекомендаций как студентам, так и преподавателям для оптимизации учебного процесса, а также помощи администрации учебных отделов вузов в виде корректировки учебных планов, если того потребует ситуация.

Цель исследования

Целью работы является создание базиса для разработки программно-математического обеспечения и автоматизации бизнес-процессов учебного отдела, в частности, Института ИТ МТУ МИРЭА.

Для достижения поставленной цели была сформулирована исследовательская гипотеза, что образовательные данные содержат сведения, которые позволят управлять процессом обучения и строить предиктивные модели.

Для проверки исследовательской гипотезы были поставлены исследовательские вопросы (ИВ):

ИВ: выявление источников образовательных данных и их классификация;

¹ Зорина Н.В. Интеллектуальный анализ образовательных данных: ретроспектива и перспективы развития / Н.В. Зорина, В.М. Панченко, А.Л. Зорин // Научная дискуссия: вопросы математики, физики, химии, биологии: сб. ст. по материалам LIII-LIV Международной научно-практической конференции «Научная дискуссия: вопросы математики, физики, химии, биологии». 2017. № 5-6(38). М.: Изд. «Интернаука», 2017. С. 42-52.



ИВ: сбор данных из образовательных источников, в том числе и на основе эксперимента, для углубленного изучения и осмысления взаимосвязей семантических закономерностей в образовательных данных;

ИВ: представление формального описание данных, полученных из источников, как системы порожденных данных;

ИВ: разработка проблемно-ориентированного математического и программного обеспечения.

Материалы и методы исследования

При проверке исходной гипотезы были исследованы образовательные данные, которые содержат сведения об учебной деятельности, полученные в результате экспериментальных наблюдений по репрезентативной выборке учащихся. Эти данные получены из нескольких источников, а также в результате проведения ряда экспериментов, описанных в работе [14]. Все студенты из наблюдаемой выборки обучаются на очном отделении по ИТ-направлению подготовки бакалавриата "Программная инженерия". Ограничения: данные содержат сведения о семестровой успеваемости студентов 3 курса и 4 курса, начиная с первого года обучения и результатах текущей образовательной деятельности (по результатам весеннего семестра 2016-2017 учебного года). Для ответа на ИВ1 были рассмотрены различные доступные для исследования источники образовательных данных и проведена их классификация. Для ответа на ИВ2 образовательные данные были собраны из источников, определенных в результате ответа на ИВ1. Для сбора данных использована методология системного анализа по системологии Дж. Клира [15] и методология проведения единичного эксперимента, используемая в нашем вузе [16]. В процессе работы над ответом на ИВ2 поставлен ряд единичных экспериментов и собраны данные с помощью специально разработанного для этой цели программного обеспечения - экспериментального программного комплекса для моделирования и интерпретации процессов анализа образовательных данных. Механизм работы и состав программного комплекса подробно описан в работе [16].

Для ответа на ИВ3 для формального описания полученных данных использовано

представление на основе математического аппарата теории множеств и алгебраических структур.

Для ответа на вопрос ИВ4 проведен анализ инструментальных средств разработки и анализа данных, обоснован выбор и определен инструментарий по реализации математического и программного обеспечения, выбраны язык, среда программирования и технология разработки. Реализованы обработка и анализа образовательных данных в виде соответствующего программного обеспечения.

Полученные результаты

ИВ1: Определение источников образовательных данных и их классификация. Результаты образовательной деятельности содержат сведения об обучении по выборке из исходных данных обучаемых Института ИТ МТУ МИРЭА. Эти сведения представляют из себя образовательные данные, которые можно разделить на априорные (ретроспективные) и апостериорные (текущие). Все данные получены из общедоступных источников, некоторые из них - в результате ряда поставленных экспериментов. Данные отражают временные ряды значимых для исследования характеристик. Выявлено, что в настоящее время образовательные данные накапливаются в информационных системах (ИС) учебных отделов деканатов в электронном и бумажном виде, требующем машинной обработки. Данные о текущем состоянии успеваемости обучаемых учитывает каждый преподаватель по каждому читаемому курсу и по каждой отдельной учебной группе студентов. Эти данные содержат следующие сведения: результаты посещаемости студентами аудиторных занятий (лекций, практических и лабораторных работ, семинаров), текущий прогресс в выполнении заданий по разделам курса (могут быть оценки за контрольные мероприятия) и результаты выполнения самостоятельной работы студентов. В каждый отдельно взятый момент времени в течении семестра каждый преподаватель накапливает сведения, которые отражают текущую успеваемость обучающихся. Накапливаемые данные содержат сведения о текущей успеваемости и никак не отражают сведения о предыдущих результатах обучения (предыстории). Все эти данные накапливаются



исключительно преподавателями, и как правило за редким исключением, в бумажном виде, и никак не обрабатываются автоматизированными системами структурных подразделений учебных заведений. Журналы учебных групп, контролируемые учебной частью, содержат сведения только о контроле посещаемости. Сведения, накопленные в ИС учебной части, наоборот, представляют из себя информацию о предыдущих результатах обучения студентов и дублируют данные зачетных книжек в электронном виде

Также для ответа на ИВ1 была проведена классификация выявленных источников образовательной информации. Выявлено, что классифицировать источники информации в виде образовательных данных можно по следующим признакам:

1. По отношению к объекту исследования.

Информация бывает внутренней и внешней. Система внутренней информации – это данные, используемые для статистического, оперативного учета, аналитики, а также формирования отчетности, характеризующей результаты обучения студентов в данном конкретном вузе. Система внешней информации – это статистические данные, публикуемые учебными заведениями в открытом доступе;

2. По отношению к предмету исследования. Информация делится на основную и вспомогательную, необходимую для более полной характеристики изучаемой предметной области. В качестве основной информации выступают образовательные данные, получаемые из зачетки и ИС учебной части деканата;

3. По периодичности поступления информации. Образовательные данные подразделяются на регулярную и эпизодическую, в зависимости от формы контроля. Итоговый контроль всегда проводится в установленные сроки – в периоды проведения зачетной и экзаменационной сессий. К источникам регулярной информации относятся семестровые данные. Эпизодическая информация формируется по мере необходимости, например, студент выполняет самостоятельную работу в определенном удобном ему темпе в заданные сроки.

В таблице 1 представлена классификация использованных в работе источников образовательных данных. Данные, представленные в четвертой строке таблицы, были получены в результате поставленного эксперимента.

Таблица 1. Классификация источников образовательных данных

№ п.п	Источники образовательных данных	Вид информации	Тип информации	Безбумажная обработка Да/Нет	Ответственный за накопление данных
	Зачетная книжка	Сведения о сданных экзаменах, зачетах, курсовых	информативный	Нет	Преподаватель
	Журнал группы	Сведения о посещаемости	контроль посещаемости	Нет	Староста группы
	Модули ИС учебной части	Сведения о не сданных и сданных экзаменах, зачетах, курсовых	контроль успеваемости	Да	Учебная часть
	Записи преподавателя	Сведения о посещаемости аудиторных занятий	текущий контроль успеваемости	Нет	Преподаватель
		Сведения о выполнении заданий на практические и л/р	текущий контроль выполнения	Нет	Преподаватель
		Сведения о выполнении самостоятельной работы	текущий контроль выполнения	Нет	Преподаватель
		Сведения о выполнении заданий на зачете и экзамене	промежуточный и итоговый контроль	Нет	Преподаватель



ИБ2: Сбор данных из образовательных источников, в том числе и на основе эксперимента для углубленного изучения и осмысления взаимосвязей, выявление закономерностей в образовательных данных.

На этом этапе были собраны образовательные данные из источников, определенных на этапе решения ИВ1. Данные, отражающие результаты посеместровой образовательной деятельности изучаемой выборки обучаемых, были получены в электронном виде в результате экспорта данных в таблицы Excel из модуля “Сессия” системы ТАНДЕМ. Университет МТУ МИРЭА Института ИТ в ручном режиме. Данные собирались в виде xls файлов сводной ведомости по группам. Сводная ведомость содержит следующую информацию: фио студента, шифр зачетки, наименование предметов, оценки по экзаменам по пятибалльной шкале, кодируемой в виде латинских букв (У, Х, О), и сведения по зачетам (З/Н). Данные о текущей успеваемости были собраны: 1) с помощью ведения учетной ведомости преподавателем посещения занятий и мониторинга выполнения заданий в течении семестра в виде таблиц Excel; 2) в виде данных, содержащих сведения о самостоятельной работе студентов, получены с помощью разработанного Экспериментального программного комплекса для моделирования и интерпретации образовательных данных с выгрузкой результатов в формате csv в автоматическом режиме.

Данные, собранные в формате электронных таблиц, представляют из себя удобный вид для последующей машинной обработки, Их легко

конвертировать в формат csv. Формат csv был выбран в качестве формата представления данных не случайно. Его удобно использовать в дальнейшем для проведения интеллектуального анализа данных. Например, этот формат используется в качестве формата входных данных для инструментального средства в составе платформы KNIME. Вид полученных в результате проведения экспериментов данных представлен на рисунках 2 и 7 в работе [17].

При выполнении исследовательской работы учитывалось, что модель обучения изменилась за последние пять лет в связи с активным внедрением цифровых способов связи и интерактивного взаимодействия. Был проведен анализ бизнес-процессов, отражающих взаимодействие учащихся между собой и взаимодействия учащегося и преподавателя в процессе обучения. Полученные результаты представлены в виде описаний бизнес-процессов в нотации bpmn на рисунках 1 и 2.

В соответствии с действующим стандартом IEEE P1484.1/D8[18] и на основе описанных бизнес-процессов предложена улучшенная архитектура образовательной системы (Architecture and Reference Model) представленная на рисунке 3.

При создании системы «Улучшенная модель архитектуры» демонстрируются на рис.3 все аспекты “совместного обучения”. Данная архитектура отражает многосторонние связи между обучаемыми и компонентами системы, а также описывает реальную модель “коллаборации” обучаемых и преподавателей, которая в настоящий момент существует.

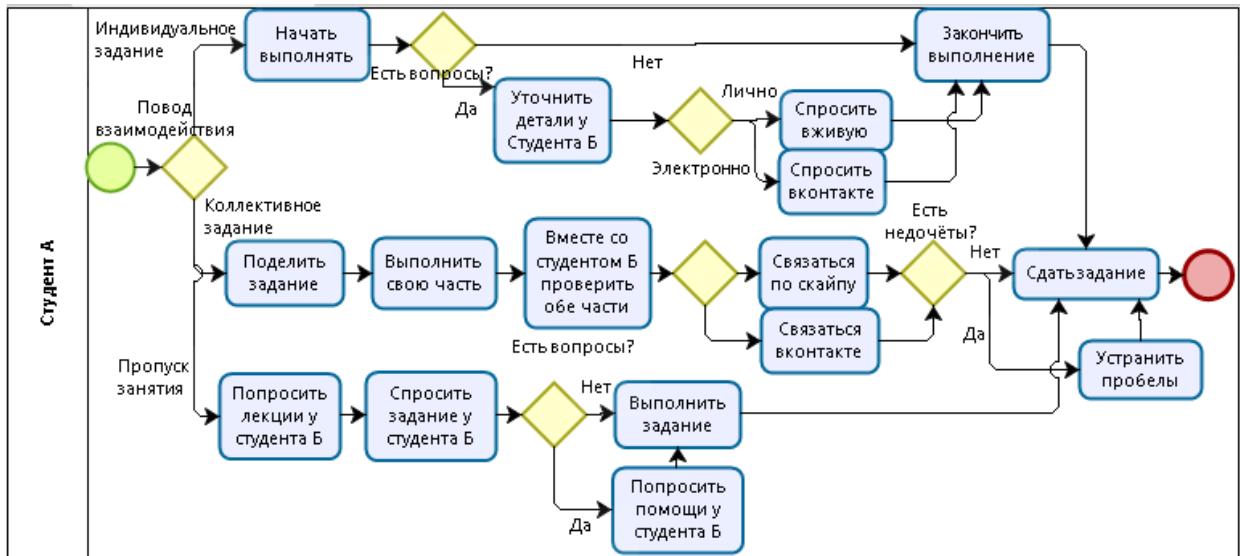


Рисунок 1. – Бизнес-процесс “обучаемый- обучаемый”

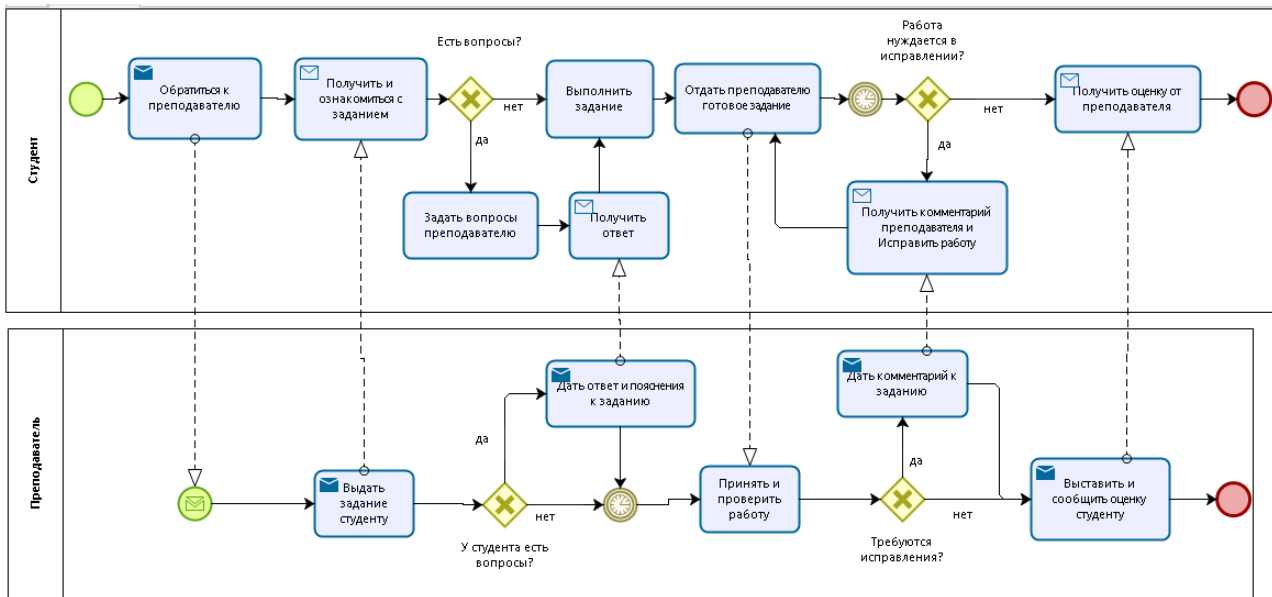


Рисунок 2. – Бизнес-процесс “обучаемый-преподаватель”

ИВЗ: Представление формального описание данных, полученных из источников в виде системы порожденных данных. Собранные на предыдущем этапе решения задач сбора образовательные данные (до “парсинга” в формат csv) представляют из себя двумерные таблицы в виде строк и столбцов, связывающие отношениями вид – свойство.

Каждая строка таблицы содержит сведения о результатах учебной деятельности обучаемых в виде оценки по пятибалльной системе, фактически этот параметр может принимать значение {3,4,5}, каждый столбец таблицы содержит сведения по отдельно взятому предмету по всем обучаемым.

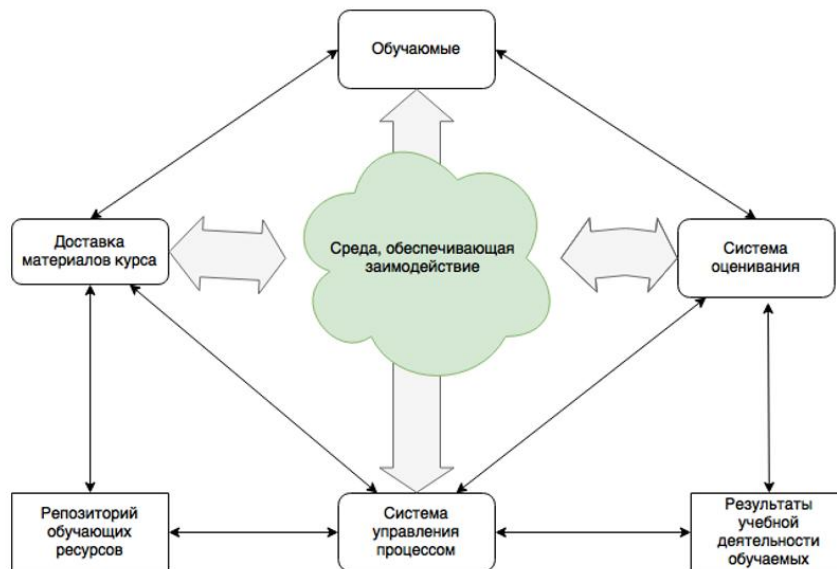


Рисунок 3. - Улучшенная модель архитектуры LMS

Обозначим множество всех обучаемых в выборке как A , а все изучаемые ими предметы как B . Пусть мощность множества A будет равняться $A = |n|$, где n -количество обучаемых в выборке A . Мощность множества B будет равняться $B = |m|$, где m -количество всех предметов, изучаемых студентами. Тогда каждый элемент в двумерной таблице будет соответствовать декартову произведению множеств A и B . Таким образом можно однозначно определить каждый элемент в таблице, как элемент, соответствующий определенному множеству результатов. Согласно базовой теореме, любая группа, являющаяся аффинным алгебраическим многообразием, допускает точное конечномерное представление, то есть ее можно задать группой матриц с элементами в поле k , заданной полиномиальными уравнениями с коэффициентами в k -пространстве чисел. Таким образом, имеем заданную алгебраическую группу:

$$G_1: A \times B \rightarrow U,$$

Здесь G_1 , означает группу, задающую отображение множества A на множество B .

$$G_2: A \times K_d \rightarrow X,$$

A в G_2 будет означать группу, задающую отображение множества A на систему

порожденных данных, где K_d - коэффициент доверия, рассчитываемый в соответствии с моделью посещаемости занятий обучаемыми. Фрагмент таблицы с данными используемой модели посещаемости группы представлен на рисунке 4.

Подобная модель используется и при расчёте коэффициентов доверия по другим параметрам оценки деятельности обучаемого, например, выполнению по выполнению заданий практикума ($K_{дз}$). Таким образом основные показатели образовательной деятельности обучаемого можно рассчитать на основании данных наблюдений, фиксируемых в ходе процесса обучения. Для вычисления средней оценки можно использовать интегрально общие результаты и по видам изучаемых дисциплин.

Таким образом формально вычисляется базовая оценка W_b для каждого обучаемого. Априорную базовую оценку можно представить в виде $W_b(Y;X)$, где Y - представляет собой $U_{ср.}$ (среднюю оценку), а X в виде вектора $X (K_d(x_1), K_d(x_2), \dots, K_d(x_n))$. Этот ряд системно определяемых коэффициентов оценки степени доверия к индивидуальной и групповой деятельности обучаемого можно, подобно $U_{ср.}$, обобщить в форме $K_d(X_{ср.})$.



ГР		№ учебной недели	1 н				2 н	...	16 н		17 н		Коэф. Доверия
№	Ф.И.О.	№ занятия	1	2	3	4		31	32	33	34	Кд	
1	Студент 1	Посещение	0	0	0	1		0	0	0	1	0,06	
2	Студент 2		0	0	0	0		0	1	0	1	0,06	
3	Студент 3		0	1	0	1		0	0	1	1	0,12	
4	Студент 4		0	0	0	0		1	0	1	1	0,09	
5	Студент 5		0	1	1	1		1	1	0	1	0,18	
6	Студент 6		0	0	1	1		1	0	0	0	0,09	
7	Студент 7		0	0	0	1		0	0	0	1	0,06	
8	Студент 8		0	1	0	1		0	1	0	1	0,12	
9	Студент 9		1	0	1	1		0	1	0	0	0,12	
10	Студент 10		0	1	1	0	...	1	1	0	1	0,15	
11	Студент 11		1	0	0	0		0	0	1	1	0,09	
12	Студент 12		0	0	1	1		1	0	1	0	0,12	
13	Студент 13		0	1	1	0		0	0	0	1	0,09	
14	Студент 14		0	1	0	0		0	1	0	0	0,06	
15	Студент 15		0	0	0	1		0	0	0	1	0,06	
16	Студент 16		0	0	0	0		1	0	0	1	0,06	
17	Студент 17		0	0	1	1		0	1	0	1	0,12	
18	Студент 18		0	0	0	0		1	1	0	1	0,09	
19	Студент 19		0	0	1	1		0	1	0	1	0,12	
20	Студент 20		0	0	0	1		0	0	0	0	0,03	

Рис. 4. - Модель формирования коэффициента доверия по посещаемости

В качестве формулы для расчёта базовой оценки для пятибалльной шкалы оценок применяется формула $W(X;Y) = 0,25y + 1,25x + 0,5yx$, в которой $y=U_{ср.}$, $x=Kд$.

Очевидно приведение данного уравнения к виду $0,25y + (1,25 + 0,5y)x$, т.е. к форме линейного уравнения $W(X;Y)=a+bx$, где $a=0,25y$, $b=1,25 + 0,5y$.

ИВ4: К разработке математического и программного обеспечения. Состав и характер исходных данных был определен на предыдущих этапах. На данном этапе производится выбор инструментального средства среды разработки и языка программирования для разработки программного обеспечения.

При выборе инструментария для разработки программного обеспечения проанализированы следующие инструментальные средства: RapidMiner (ранее известный как YALE) [19], WEKA [20,21], ORANGE [22,23], KEEL [24], KNIME [25,26]. Каждое из этих средств имеет свои достоинства и недостатки. Аналитические платформы RapidMiner и KNIME схожи между собой по ряду характеристик. Обе позволяют конечным пользователям использовать визуальное программирование потоков выполнения анализа. RapidMiner имеет расширенный функционал при анализе неструктурированной информации и более понятный интерфейс, но лицензия для неё будет бесплатной только при использовании в исследовательских целях. При использовании в

рамках академической лицензии RapidMiner имеет ограниченный функционал. Weka – бесплатный продукт, включает большинство алгоритмов проведения анализа, но имеет неудобный пользовательский интерфейс. Платформа KNIME имеет много общего с инструментами RapidMiner и Weka, фактически, объединяя их возможности. KNIME содержит набор инструментов для очистки, анализа и визуализации данных, фактически как у RapidMiner, а также позволяет использовать те же алгоритмы, что и пакет Weka.

Knime Analytics Platform (KNIME) представляет “open source” фреймворк для анализа данных. Данный фреймворк позволяет реализовывать полный цикл анализа данных, включающий все этапы работы с данными: чтение данных из различных источников, их преобразование, фильтрацию, анализ, визуализацию и экспорт данных. В процессе исследования принципов работы с KNIME выяснилось, что особым преимуществом KNIME является его способность интегрировать данные сразу из нескольких источников (например, из CSV файлов и из текстовых документов, данных из соцсетей ответов и баз данных, содержащих демографическую информацию об юбучаемых) в рамках одного анализа. Еще одно преимущество KNIME – универсальность использования, так как фреймворк включает расширения, которые позволяют ему взаимодействовать с языками R, Python, Java и SQL. Окно с рабочим потоком KNIME представлено на рисунке 5.

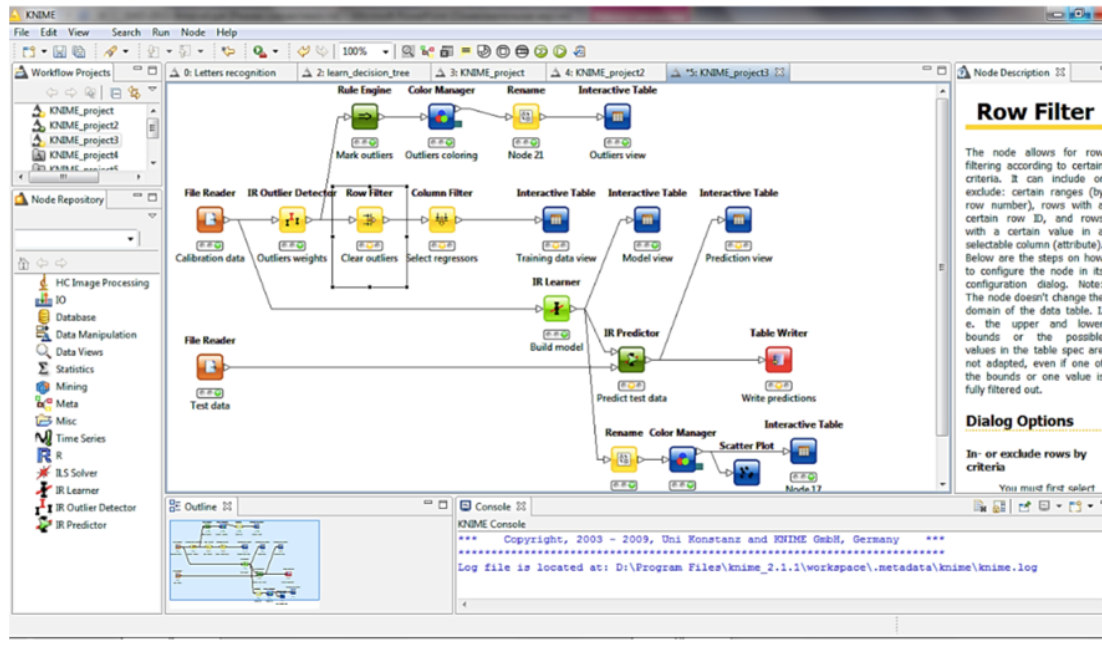


Рис. 5. - Окно с рабочим потоком KNIME

Были также проанализированы возможности использования *облачных сервисов для анализа данных и машинного обучения* Amazon Machine Learning (Amazon) и Azure Machine Learning (Microsoft). Необходимо отметить, что ряд учебных заведений в США применяет технологию Amazon Machine Learning. Этот сервис упрощает использование технологий машинного обучения и предназначен для разработчиков любого уровня квалификации. Преимуществами сервиса Amazon Machine Learning является то, что он включает в себя встроенные инструменты визуализации и специальные “визарды” для создания моделей машинного обучения (ML), конечным пользователям не нужно создавать собственный код для формирования прогнозной аналитики, а также управлять какой-либо инфраструктурой. Недостатком для российского потребителя является стоимость таких проектов и сроки окупаемости. Все вышеперечисленные инструменты для создания программного обеспечения можно использовать для проведения интеллектуального анализа данных, но по мнению авторов, наиболее подходящими для решения задачи анализа образовательных данных по совокупности параметров являются один из них – фреймворк KNIME.

В рамках задач, поставленных в настоящей

исследовательской работе, для аналитической платформы KNIME был разработан и реализован новый рабочий узел в виде плагина для KNIME в IDE Eclipse. Узел реализует задачи анализа образовательных данных в виде вычислительных процессов, описанных в работах [27,28]. Хотя фреймворк KNIME позволяет писать код на языках R или Python, но в качестве языка программирования для реализации программного обеспечения был выбран Java API. Язык программирования Java, с точки зрения авторов, является наиболее удобным для программирования модулей KNIME и органично вписывается в среду, так как сам Фреймворк написан на Java.

Заключение

Адаптация и персонафикация учебного процесса с использованием современных ИТ-технологий и технологий Big Data в настоящий момент является несомненным трендом высшего образования. Проблема унаследованных ИС отечественных вузов в том, что накопленные и накапливаемые данные о результатах образовательной деятельности учащихся в настоящее время никак не используются. Системный подход к автоматизации основных бизнес-процессов поддержки учебного процесса на основе интеллектуального анализа данных позволит использовать накопленные и накапливаемые



данные для учебной аналитики и прогнозирования результатов обучения. Описанное в статье исследование проводилось с целью формирования программно-математического обеспечения для решения описанных проблем на примере ИТ-образования отечественного вуза. Был при этом поставлен и решен ряд исследовательских вопросов. Разработаны модели вычислительных процессов и предложены средства для формирования вычислительного модуля.

Ведутся работы по дальнейшему усовершенствованию разработанной программной компоненты в качестве реализации механизма интеллектуального анализа данных на основе инструментария KNIME. Для бесшовной интеграции со средами дистанционного обучения LMS планируется также разработать обертку на основе протокола REST, которая позволит внешним объектам получать доступ к аналитическим возможностям разработанного модуля.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Келейникова В.А, Романова Е.В. К вопросу реализации ИТ-образования в рамках концепции Smart Education // Прикладная информатика. 2014. № 1(49). С. 82-90. URL: <https://elibrary.ru/item.asp?id=21223150> (дата обращения: 01.02.2018).
- [2] Bienkowski M., Feng M., Means B. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, Washington, D.C., 2012. URL: <https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf> (дата обращения: 01.02.2018).
- [3] Calvet Liñán, L., Juan Pérez, Á.A. Educational Data Mining and Learning Analytics: differences, similarities, and time evolution // RUSC. Universities and Knowledge Society Journal. 2015. Vol. 12, no. 3. Pp. 98-112. DOI: <http://dx.doi.org/10.7238/rusc.v12i3.2515>
- [4] Romero C., Ventura S. Data mining in education // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2013. Vol. 3, no. 1. Pp. 12-27. DOI: <https://doi.org/10.1002/widm.1075>
- [5] Romero C., Ventura S., Pechenizkiy M., Baker R.S.J.D. (Eds.) Handbook of Educational Data Mining. Boca Ratón, FL: CRC Press. 2010. 503 p.
- [6] Nithya P., Umamaheswari B., Umadevi A. A Survey on Educational Data Mining in Field of Education // International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). 2016. Vol. 5, issue 1. Pp. 69-78. URL: <http://ijarcet.org/wp-content/uploads/IJARCET-VOL-5-ISSUE-1-69-78.pdf> (дата обращения: 01.02.2018).
- [7] Jasvinder K. A Comprehensive Study of Educational Data Mining // International Journal of Electrical Electronics & Computer Science Engineering. Special Issue – TeLMISR 2015. 2015. Pp. 58-63. URL: <http://www.ijeecse.com/T110.pdf> (дата обращения: 01.02.2018).
- [8] Next-term student grade prediction / M. Sweeney, J. Lester, H. Rangwala // Proceedings of 2015 IEEE International Conference on Big Data (Big Data) (BIG DATA '15). IEEE Computer Society, Washington, DC, USA, 2015. Pp. 970-975. DOI: <https://doi.org/10.1109/BigData.2015.7363847>
- [9] Peña-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works // Expert Systems with Applications. 2014. Vol. 41, issue 4, Part 1. Pp. 1432-1462. DOI: <https://doi.org/10.1016/j.eswa.2013.08.042>
- [10] Knight S., Shum S.B. Theory and Learning Analytics. In Lang C., Siemens G., Wise A.F., Gaevic D., editors, The Handbook of Learning Analytics. 2017. Pp. 17-22. Society for Learning Analytics Research (SoLAR), Alberta, Canada, 1 edition. DOI: <https://doi.org/10.18608/hla17>
- [11] Shaffer D.W., Collier W., Ruis A.R. A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data // Journal of Learning Analytics. 2016. Vol. 3, no. 3. Pp. 9-45. DOI: <http://dx.doi.org/10.18608/jla.2016.33.3>
- [12] Romero C., López M.-I., Luna J.-M., Ventura S. Predicting students' final performance from participation in on-line discussion forums // Computers & Education. 2013. Vol. 68. Pp. 458-472. DOI: <https://doi.org/10.1016/j.compedu.2013.06.009>
- [13] Kavitha G., Raj L. Educational Data Mining and Learning Analytics - Educational Assistance for Teaching and Learning // Internal Journal of Computer and Organization Trends (IJCOT). 2017. Vol. 7, issue 2. Pp. 21-25. DOI: <https://doi.org/10.14445/22492593/IJCOT-V41P304>
- [14] Зорина Н.В. Интеллектуальный анализ данных результатов единичных экспериментов для оценки социальных объектов наблюдения: Дис. ... магистра тех. наук. Москва: МТУ МИРЭА, 2017, 78 с.
- [15] Клир Дж. Системология. Автоматизация решения системных задач / Дж. Клир; Перевод с англ. М.А. Зуева; Под ред. А.И. Горлина. М.: Радио и связь, 1990. 544 с.
- [16] Панченко В.М. Компьютерные технологии и системы обучения. Технология разделяемых единиц контента в системе программ единичных экспериментов. Часть 2. М.: МИРЭА, 2008. 156 с.
- [17] Зорина Н.В., Панченко В.М. Экспериментальный программный комплекс для моделирования и интерпретации процессов анализа образовательных данных // Современные информационные технологии и ИТ-образование. 2017. Т. 13, № 4. С. 207-215. DOI: <https://doi.org/10.25559/SITITO.2017.4.483>
- [18] IEEE 1484.12.3. Standard for Learning Technology – Extensible Markup Language (XML) Schema Definition Language Binding for Learning Object Metadata. URL: <http://standards.ieee.org/findstds/standard/1484.12.3-2005.html> (дата обращения: 01.02.2018).
- [19] Rapid Mining Data Mining. URL: <http://rapidminer.com/resources> (дата обращения: 01.02.2018).



- [20] Waikato Environment for Knowledge Analysis. URL: <https://www.cs.waikato.ac.nz/ml/weka> (дата обращения: 01.02.2018).
- [21] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H. The WEKA data mining software: an update // *ACM SIGKDD Explor. Newsl.* 2009. Vol. 11, issue 1. Pp. 10-18. DOI: <http://dx.doi.org/10.1145/1656274.1656278>
- [22] Orange Data Mining. URL: <https://orange.biolab.si> (дата обращения: 01.02.2018).
- [23] Demšar J., Curk T., Erjavec A., Gorup T., Hocevar T., Milutinovic M., Možina M., Polajnar M., Toplak M., Staric A., Štajdohar M., Umek L., Žagar L., Žbontar J., Žitnik M., Zupan B. Orange: Data mining toolbox in python // *Journal of Machine Learning Research.* 2013. Vol. 14(Aug). Pp. 2349–2353. URL: <http://www.jmlr.org/papers/volume14/demsar13a/demsar13a.pdf> (дата обращения: 01.02.2018).
- [24] Narwal M., Mittal P. Knowledge Extraction based on Evolutionary Learning (KEEL): Analysis of Development Method, Genetic Fuzzy System // *International Journal of Computer Applications & Information Technology.* 2012. Vol. 1, no. 1. Pp. 22-25. URL: <http://ijcait.com/IJCAIT/117.pdf> (дата обращения: 01.02.2018).
- [25] Berthold M.R. et al. KNIME: The Konstanz Information Miner. In: Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization.* Springer, Berlin, Heidelberg, 2008. Pp. 319-326. DOI: https://doi.org/10.1007/978-3-540-78246-9_38
- [26] Knime Analytics Platform. URL: <https://www.knime.com/resources> (дата обращения: 01.02.2018).
- [27] Зорина Н.В., Панченко В.М. Описание объекта моделирования для постановки эксперимента по оцениванию результатов деятельности обучаемого по данным программ единичного эксперимента // *Интернаука.* 2017. № 11(15). Часть 1. С. 25-27.
- [28] Зорина Н.В. Интеллектуальный анализ образовательных данных: ретроспектива и перспективы развития / Н.В. Зорина, В.М. Панченко, А.Л. Зорин // *Научная дискуссия: вопросы математики, физики, химии, биологии: сб. ст. по материалам LIII-LIV Международной научно-практической конференции «Научная дискуссия: вопросы математики, физики, химии, биологии».* 2017. № 5-6(38). М., Изд. «Интернаука», 2017. С. 42-52.

Поступила 10.12.2017; принята к публикации 01.02.2018; опубликована онлайн 30.03.2018.

REFERENCES

- [1] Keleynikova V.A., Romanova E.V. Questions of realization of IT-education using the concept of smart-education. *Applied informatics.* 2014; 1(49):82-90. Available at: <https://elibrary.ru/item.asp?id=21223150> (accessed 01.02.2018). (In Russian)
- [2] Bienkowski M., Feng M., Means B. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, Washington, D.C., 2012. Available at: <https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf> (accessed 01.02.2018).
- [3] Calvet Liñán, L., Juan Pérez, Á.A. Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *RUSC. Universities and Knowledge Society Journal.* 2015; 12(3):98-112. DOI: <http://dx.doi.org/10.7238/rusc.v12i3.2515>
- [4] Romero C., Ventura S. Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.* 2013; 3(1):12–27. DOI: <https://doi.org/10.1002/widm.1075>
- [5] Romero C., Ventura S., Pechenizkiy M., Baker R.S.J.D. (Eds.) *Handbook of Educational Data Mining.* Boca Raton, FL: CRC Press. 2010. 503 p.
- [6] Nithya P., Umamaheswari B., Umadevi A. A Survey on Educational Data Mining in Field of Education. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET).* 2016; 5(1):69-78. Available at: <http://ijarcet.org/wp-content/uploads/IJARCET-VOL-5-ISSUE-1-69-78.pdf> (accessed 01.02.2018).
- [7] Kumar J. A Comprehensive Study of Educational Data Mining. *International Journal of Electrical Electronics & Computer Science Engineering. Special Issue - TeLMISR 2015.* 2015. p. 58-63. Available at: <http://www.ijeecse.com/T110.pdf> (accessed 01.02.2018).
- [8] Sweeney M., Lester J., Rangwala H. Next-term student grade prediction. *Proceedings of 2015 IEEE International Conference on Big Data (Big Data) (BIG DATA '15).* IEEE Computer Society, Washington, DC, USA, 2015. p. 970-975. DOI: <https://doi.org/10.1109/BigData.2015.7363847>
- [9] Peña-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications.* 2014; 41(4-1):1432-1462. DOI: <https://doi.org/10.1016/j.eswa.2013.08.042>
- [10] Knight S., Shum S.B. Theory and Learning Analytics. In Lang C., Siemens G., Wise A.F., Gaevic D., editors, *The Handbook of Learning Analytics.* 2017. p. 17–22. Society for Learning Analytics Research (SoLAR), Alberta, Canada, 1 edition. DOI: <https://doi.org/10.18608/hla17>
- [11] Shaffer D.W., Collier W., Ruis A.R. A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics.* 2016; 3(3):9-45. DOI: <http://dx.doi.org/10.18608/jla.2016.33.3>
- [12] Romero C., López M.-I., Luna J.-M., Ventura S. Predicting students' final performance from participation in on-line discussion forums. *Computers & Education.* 2013; 68:458-472. DOI: <https://doi.org/10.1016/j.compedu.2013.06.009>
- [13] Kavitha G., Raj L. Educational Data Mining and Learning Analytics - Educational Assistance for Teaching and Learning. *International Journal of Computer and Organization Trends (IJCOT).* 2017; 7(2):21-25. DOI: <https://doi.org/10.14445/22492593/IJCOT-V41P304>
- [14] Zorina N.V. Intellectual analysis of data from the results of individual experiments to assess social objects of observation: Dis. ... Master of Tech. sciences. Moscow: MTU MIREA, 2017, 78 p. (In Russian)
- [15] Klir D.J. *Systemology. Automation of system task solution.* M.: Radio and Communication, 1990. 544 p. (In Russian)
- [16] Panchenko V.M. Computer technologies and training systems. The technology of shared content units in the system of programs of single experiments. Part 2. M.: MIREA, 2008. 156 p. (In Russian)
- [17] Zorina N.V., Panchenko V.M. Experimental software for modeling and interpreting educational data analysis processes. *Modern Information Technology and IT-education.* 2017; 13(4):207-215. (in Russian) DOI: <https://doi.org/10.25559/SITITO.2017.4.483>



- [18] IEEE 1484.12.3. Standard for Learning Technology – Extensible Markup Language (XML) Schema Definition Language Binding for Learning Object Metadata. Available at: <http://standards.ieee.org/findstds/standard/1484.12.3-2005.html> (accessed 01.02.2018).
- [19] Rapid Mining Data Mining. Available at: <http://rapidminer.com/resources> (accessed 01.02.2018).
- [20] Waikato Environment for Knowledge Analysis. Available at: <https://www.cs.waikato.ac.nz/ml/weka> (accessed 01.02.2018).
- [21] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H. The WEKA data mining software: an update. *ACM SIGKDD Explor. Newsl.* 2009; 11(1):10-18. DOI: <http://dx.doi.org/10.1145/1656274.1656278>
- [22] Orange Data Mining. Available at: <https://orange.biolab.si> (accessed 01.02.2018).
- [23] Demšar J., Curk T., Erjavec A., Gorup Ćrt., Hocevar T., Milutinovic M., Možina M., Polajnar M., Toplak M., Staric A., Štajdohar M., Umek L., Žagar L., Žbontar J., Žitnik M., Zupan B. Orange: Data mining toolbox in python. *Journal of Machine Learning Research.* 2013; 14(Aug):2349–2353. Available at: <http://www.jmlr.org/papers/volume14/demsar13a/demsar13a.pdf> (accessed 01.02.2018).
- [24] Narwal M., Mittal P. Knowledge Extraction based on Evolutionary Learning (KEEL): Analysis of Development Method, Genetic Fuzzy System. *International Journal of Computer Applications & Information Technology.* 2012; 1(1):22-25. Available at: <http://ijcait.com/IJCAIT/117.pdf> (accessed 01.02.2018).
- [25] Berthold M.R. et al. KNIME: The Konstanz Information Miner. In: Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization.* Springer, Berlin, Heidelberg, 2008. p. 319-326. DOI: https://doi.org/10.1007/978-3-540-78246-9_38
- [26] Knime Analytics Platform. Available at: URL: <https://www.knime.com/resources> (accessed 01.02.2018).
- [27] Zorina N.V., Panchenko V.M. Description of the simulation object for setting up an experiment to evaluate the results of the trainee's activity according to the programs of a single experiment. *Internauka.* 2017; no. 11-1(15):25-27. (In Russian)
- [28] Zorina N.V. Intellectual analysis of educational data: a retrospective and development prospects / N.V. Zorina, V.M. Panchenko, A.L. Zorin // Scientific discussion: Mathematics, Physics, Chemistry, Biology: coll. Art. on the materials of the LIII-LIV International Scientific and Practical Conference "Scientific discussion: Mathematics, Physics, Chemistry, Biology". 2017. No. 5-6(38). M., ed. "Internauka", 2017. Pp. 42-52. (In Russian)

Submitted 10.12.2018; Revised 01.02.2018; Published 30.03.2018.

About the authors:

Natalya V. Zorina, senior lecturer, Department of Instrumental and applied software, MIREA - Russian Technological University (78 Vernadsky Av., Moscow 119454, Russia); ORCID: <http://orcid.org/0000-0001-7893-9227>, zorina@mirea.ru

Victor M. Panchenko, candidate of technical Sciences, Professor, Department of Instrumental and applied software, MIREA - Russian Technological University (78 Vernadsky Av., Moscow 119454, Russia); ORCID: <http://orcid.org/0000-0002-7038-878X>, pvm36@yandex.ru



This is an open access article distributed under the Creative Commons Attribution License which unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).